

UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA
PROGRAMA DE EDUCAÇÃO CONTINUADA EM ENGENHARIA
ESPECIALIZAÇÃO EM INTELIGÊNCIA ARTIFICIAL

Gabriel Elias Mandanda

Previsibilidade das Tempestades Elétricas na Bacia do Prata

São Paulo
2023

GABRIEL ELIAS MANDANDA

Previsibilidade das Tempestades Elétricas na Bacia do Prata

— Versão Original —

Monografia apresentada ao Programa de Educação Continuada em Engenharia da Escola Politécnica da Universidade de São Paulo como parte dos requisitos para conclusão do curso de Especialização em Inteligência Artificial. Orientador: Prof. Dr. Marlon

Sproesser Mathias

São Paulo
2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo-na-publicação

Mandanda, Gabriel Elías

Previsibilidade das Tempestades Elétricas na Bacia do Prata/ G. E.Mandanda – São Paulo, 2023.

150p.

Monografia (Especialização em Inteligência Artificial) – Escola Politécnica da Universidade de São Paulo. PECE – Programa de Educação Continuada em Engenharia.

Tempestades elétricas, Rede convolucional 3D, Previsão de curto prazo.

I. Universidade de São Paulo. Escola Politécnica. PECE – Programa de Educação Continuada em Engenharia. II.t.

Aos meus pais, Elias e Helena.

Agradecimentos

Em primeiro agradeço a Deus pela vida e saúde e por resplandecer a sua luz nos meus caminhos.

Ao meu orientador Prof. Dr. Marlon Sproesser Mathias pela disponibilidade durante a realização deste trabalho. Estou grato pela paciência e ensinamentos.

À Profa. Dra. Maria Assunção Faus da Silva Dias pelas contribuições no melhoramento deste trabalho.

À Profa. Dra. Larissa Driemeier, coordenadora do curso, e ao coordenador Thiago de Castro Martins, pelos valiosos ensinamentos.

À minha esposa Suely e a minha filha Yasmin por me apoiarem desde o início até o fim do curso.

Aos colegas e amigos, Lucas Cabral, Rafael Costa e Wesley Ferreira Maia pelos estudos que compartilhamos juntos.

Aos funcionários do Programa de Educação Continuada da Poli-USP pela assistência excepcional às minhas dúvidas. Não posso deixar de reconhecer o excelente trabalho realizado por Agnaldo e Paula.

Ao Programa de Educação Continuada da Poli-USP (PECE) pela bolsa de estudo que me concederam para a realização do curso. Os conhecimentos adquiridos durante o curso foram fundamentais para minha pesquisa de doutorado.

Qualquer nova tecnologia
gradualmente cria um novo
ambiente para o ser humano.

— *Marshall McLuhan*

Sumário

Sumário • *v*

Resumo • *vi*

Abstract • *vii*

Lista de Figuras • *viii*

1 Introdução • *1*

1.1 Motivação • *2*

2 Revisão da literatura • *3*

2.1 Tempestades Elétricas • *3*

2.2 Aprendizado de Máquina • *4*

2.2.1 Análise de Componentes Principais • *4*

2.2.2 Análise de Cluster • *4*

2.2.3 Rede Neural Convolucional • *5*

2.3 Estado da Arte do Uso dos Modelos Supervisionados na Previsão das Tempestades Severas • *6*

3 Dados e Metodologia • *8*

3.1 Dados • *8*

3.1.1 Descrição dos Parâmetros Atmosféricos Usados na Pesquisa • *9*

3.2 Metodologia • *10*

4 Resultados e Discussão • *12*

4.1 Análise da Distribuição das Descargas Elétricas na América do Sul • *12*

4.1.1 Análise de Cluster Baseado nas Médias Mensais das Descargas Elétricas na América do Sul • *13*

4.1.2 Análise de Cluster Baseado no Ciclo diurno das Descargas Elétricas na América do Sul • *15*

4.1.3 Análise de Componentes Principais • *18*

4.2 Previsão das Descargas Elétricas na Bacia do Prata • *19*

5 Conclusão • *22*

5.1 Limitações e Trabalhos Futuros • *22*

Referências • *23*

Resumo

MANDANDA, G. E. *Previsibilidade das Tempestades Elétricas na Bacia do Prata*. 2023. Monografia (Especialização em Inteligência Artificial) – Escola Politécnica da Universidade de São Paulo. PECE – Programa de Educação Continuada em Engenharia. Universidade de São Paulo, São Paulo, 2023.

A América do Sul é propensa à ocorrência de tempestades elétricas. Prever essas tempestades com antecedência torna-se crucial para reduzir seus impactos negativos. Nesse contexto, este trabalho visa prever as descargas elétricas na Bacia do Prata (região que engloba Argentina, Paraguai, Uruguai e o Sul do Brasil) com 1 hora de antecedência. Para atingir esse objetivo, foram utilizados os dados horários do sensor GLM (do inglês, Geostationary Lightning Mapper) do satélite GOES 16 (GOES do inglês, Geostationary Operational Environmental Satellite) e os dados de reanálise do Era-5 ao longo de três anos (de janeiro de 2018 a dezembro de 2020). O conjunto de dados possui uma resolução temporal de 1 hora e resolução espacial de 25 km. A metodologia envolve a análise das distribuições mensais e horárias das descargas elétricas na América do Sul, utilizando o algoritmo k-means para identificar e agrupar regiões com padrões semelhantes. Em seguida, para prever as descargas elétricas na Bacia do Prata, desenvolveu-se uma arquitetura com uma rede convolucional 3D e uma camada LSTM (do inglês, Long Short-Term Memory), projetada para extrair as características espaciais e temporais dos dados. Os resultados obtidos apresentam um erro absoluto médio de 2.4 raios por hora, o que representa um bom desempenho do modelo no mapeamento das áreas com maior densidade das descargas elétricas.

Palavras-chave: Tempestades elétricas, Rede convolucional 3D, Previsão de curto prazo.

Abstract

MANDANDA, G. E. *Predictability of thunderstorms in the River Plate Basin*. 2023. Monografia (Especialização em Inteligência Artificial) – Escola Politécnica da Universidade de São Paulo. PECE – Programa de Educação Continuada em Engenharia. University of São Paulo, São Paulo, Brazil. 2023.

South America is vulnerable to thunderstorms. Forecasting these storms is crucial to reducing their negative effects. In this context, the aim of this work is the one-hour forecast of electrical discharges in the La Plata basin (region that includes Argentina, Paraguay, Uruguay, and Southern Brazil). In order to achieve this goal, hourly data from the GLM sensor of the GOES 16 satellite and data from the Era-5 reanalysis over a three-year period (from January 2018 to December 2020) have been used. The dataset has 1 hour temporal resolution and 25 km spatial resolution. The methodology involves analysing the monthly and hourly distribution of electrical discharges in South America, using the k-means algorithm to identify and group regions with similar patterns. Then, an architecture with a 3D convolutional network and an LSTM layer was developed to extract the spatial and temporal features of the data to predict the electrical discharges in the La Plata Basin. The results obtained showed an average absolute error of 2.4 lightning strikes per hour, representing a good performance of the model in mapping the areas with the highest density of lightning strikes.

Keywords: Thunderstorms, 3D convolutional network, Nowcasting.

Lista de Figuras

2.1	Esquema da estrutura básica de carga na região convectiva de uma tempestade. Fonte: Stolzenburg, Rust e Marshall (1998), acessada no site da NOAA em 2023 no seguinte endereço: https://www.nssl.noaa.gov/education/svrwx101/lightning/types/	3
4.1	Método de cotovelo. Onde WCSS é a distância total dos pontos dos dados em relação aos seus centroides.	12
4.2	Análise de cluster na América do Sul baseando-se no ciclo anual. Onde cluster 1 (vermelho), cluster 2 (azul), cluster 3 (verde), cluster 4 (roxo).	13
4.3	Ciclo anual das descargas elétricas em cada região(Clusters).	14
4.4	Análise de cluster na América do Sul (baseando-se na variação diurna das descargas elétricas). Onde cluster 1 (vermelho), cluster 2 (azul), cluster 3 (verde), cluster 4 (roxo).	16
4.5	Ciclo diurno (horário local) das descargas elétricas em cada região (clusters).	17
4.6	Análise de cluster na América do Sul, baseado nos 5 componentes principais.	18
4.7	Função de perda do (a) treinamento e (b) da validação.	19
4.8	Acumulado das descargas elétricas do mês de Setembro do ano 2020.	20
4.9	Número total das descargas elétricas em cada ponto de grade. Onde (a) descargas elétricas previstas e (b) observadas.	20
4.10	Acumulado horário das descargas elétricas. Onde a linha laranja são descargas elétricas previstas e a linha azul são as descargas elétricas observadas.	21
4.11	Métricas estatísticas	21

Introdução

As tempestades elétricas são nuvens que produzem pelo menos uma descarga elétrica. A ocorrência de descargas elétricas causa danos significativos à saúde e à vida na sociedade, bem como custos monetários consideráveis (Holle 2016). Isso ocorre porque as descargas elétricas podem provocar incêndios, danos em infraestruturas elétricas e, em casos extremos, pode resultar em perdas de vidas humanas.

A América do Sul é uma das regiões do mundo onde as tempestades elétricas ocorrem com maior frequência (Cecil, Goodman et al. 2005 e Albrecht et al. 2016). Para Cecil, Goodman et al. (2005) e Zipser et al. (2006) a região que abrange Argentina, Paraguai, Uruguai e o Sul do Brasil (região designada como Bacia do Prata) é um local com potencial para ocorrência das tempestades mais intensas do mundo. Peterson (2021) documentou na sua pesquisa a ocorrência de descargas elétricas que estabeleceram um novo recorde mundial, com uma extensão de aproximadamente 709 km na Bacia do Prata. Segundo o autor (Peterson 2021) essas descargas elétricas ocorreram em um ambiente de sistemas convectivos de mesoescala.

Os estudos de Anselmo (2015) e Morales Rodriguez (2019) exploraram a climatologia das descargas elétricas na América do Sul, enfatizando ciclos diurnos e anuais em todas as suas regiões e destacando variações de distribuição de acordo com a latitude. O presente trabalho também analisou os ciclos diurnos e anuais das descargas elétricas na América do Sul aplicando a técnica de agrupamento k-means. A técnica k-means permitiu identificar as sub-regiões na América do Sul com características semelhantes de descargas elétricas. Através dessas sub-divisões tornou-se possível a definição da área de estudo, denominada por Bacia do Prata.

A dinâmica das tempestades elétricas é complexa e a sua previsão se torna mais desafiante. Portanto, o uso dos modelos de aprendizados de máquina têm sido documentado como uma das ferramenta importante para auxiliar na previsão das tempestades convectivas. Manzato (2013) desenvolveu uma rede neural para prever a ocorrência de

granizo na Itália. Gagne et al. (2017) documentaram que as previsões de aprendizado de máquinas apresentam maior confiabilidade na previsão de granizo severo. Bao et al. (2022) fizeram um sistema de previsão de relâmpagos usando técnicas de aprendizado profundo e os seus resultados experimentais demonstraram a eficiência deste modelo de inteligência artificial na emissão de alertas de aproximação de relâmpagos em curto prazo. Na América do Sul, especificamente no Brasil, Perez e Silva Dias (2018) criaram rede neural profunda para aprimorar a previsão de precipitação na região leste da cidade de São Paulo e o seu modelo forneceu previsões confiáveis. Seguindo esse conceito, o principal objetivo desta pesquisa é prever a ocorrência das descargas elétricas na Bacia do Prata usando redes convolucionais 3D.

1.1 Motivação

As informações meteorológicas desempenham um papel de extrema importância na sociedade, especialmente na prevenção de prejuízos e na garantia do funcionamento adequado das atividades cotidianas. Albrecht et al. (2016) e Peterson (2021) ressaltam a frequência das tempestades elétricas na Bacia do Prata. Cardoso et al. (2014) documentaram que, entre 2000 e 2009 no Brasil, em média morriam 132 pessoas por ano devido as descargas elétricas. Segundo Cardoso et al. (2014) maior parte das vítimas das descargas elétricas são os agricultores, que são expostos ao ar livre durante as suas atividades. Portanto, a precisão na identificação dos momentos em que as tempestades elétricas são mais prováveis, através de redes convolucionais, pode trazer um grande benefício para a meteorologia operacional. Isso fortalecerá a capacidade de previsão de tempestades elétricas a curto prazo e a emissão de alertas mais confiáveis. Dessa maneira, a sociedade será apoiada e as atividades em diversos setores poderão continuar funcionando de forma segura.

Revisão da literatura

Neste capítulo será apresentada a revisão bibliográfica sobre as tempestades elétricas, algoritmos de aprendizado de máquina e o estado da arte do uso dos modelos de aprendizado de máquina na previsão das tempestades.

2.1 Tempestades Elétricas

Tempestades elétricas são formadas por nuvens do tipo cumulonimbus (Cb) que geram descargas elétricas. A origem dessas descargas dentro das nuvens ocorre devido à colisão entre diferentes hidrometeoros, como graupel, cristais de gelo e granizo, permitindo a transferência de cargas elétricas. Essas descargas representam correntes elétricas que se deslocam entre centros de cargas de polaridades opostas. Existem diferentes tipos de descargas elétricas: intra-nuvem, que ocorrem inteiramente dentro de uma única nuvem de tempestade; nuvem-nuvem, que ocorrem entre duas nuvens de tempestade distintas; nuvem-ar, que se inicia na nuvem e termina fora dela; e nuvem-solo, que se origina na nuvem e atinge o solo, ou vice-versa (Reynolds, Brook e Gourley 1957).

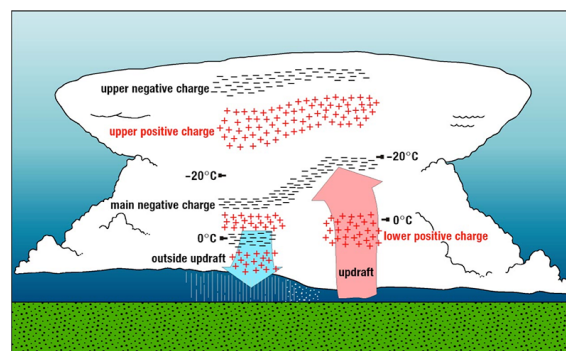


Figura 2.1: Esquema da estrutura básica de carga na região convectiva de uma tempestade. Fonte: Stolzenburg, Rust e Marshall (1998), acessada no site da NOAA em 2023 no seguinte endereço: <https://www.nssl.noaa.gov/education/svrwx101/lightning/types/>.

A pesquisa desenvolvida por Stolzenburg, Rust e Marshall (1998) baseada em medições de sondagens de campo elétrico em tempestades permitiu mostrar a distribuição das descargas elétricas em tempestades na Figura 2.1.

2.2 Aprendizado de Máquina

O Aprendizado de Máquina é uma área da Inteligência Artificial (IA) que se concentra no desenvolvimento de algoritmos e modelos que permitem que as máquinas aprendam a partir dos dados (Bishop e Nasrabadi 2006). Durante o processo de aprendizado, as máquinas identificam padrões e erros, na qual permite melhorar o seu desempenho, para a tomada de decisão sem necessidade de intervenção direta do homem. Aprendizado de máquina divide-se em três principais tipos: aprendizado supervisionado (o modelo aprende com os dados rotulados), aprendizado não supervisionado (o modelo aprende padrões sem dados rotulados) e aprendizado por reforço (o modelo aprende com o retorno das avaliações). Nesta pesquisa foram desenvolvidos modelos de aprendizado não supervisionados e supervisionados.

2.2.1 Análise de Componentes Principais

PCA (Principal Component Analysis), que significa Análise de Componentes Principais, é uma técnica de aprendizado não supervisionado eficaz e amplamente utilizada na estatística multivariada para redução da dimensionalidade. PCA emprega a decomposição de valor singular, proporcionando a melhor aproximação de baixo nível para os dados originais. Neste trabalho foi aplicada a técnica de PCA para a redução da dimensão da amostra das descargas elétricas.

2.2.2 Análise de Cluster

A Análise de Cluster é uma técnica de aprendizado não supervisionado amplamente empregada na análise exploratória de dados, sendo frequentemente utilizada em diversas áreas (Ben-Dor e Yakhini 1999 e Valli 2002) com o propósito de investigar a estrutura subjacente dos dados. Existem vários tipos de algoritmos que agrupam dados com características semelhantes. Neste trabalho, será descrito apenas o algoritmo k-means, que foi utilizado na metodologia desta pesquisa. Fotouhi e Montazeri-Gh (2013) documentaram que o agrupamento k-means requer menos poder computacional em comparação com outros métodos de análise de cluster. De acordo com Gan, C. Ma e Wu (2020), o algoritmo k-means é notavelmente rápido e fácil de ser implementado. O algoritmo k-means foi desenvolvido há mais de 50 anos por Steinhaus et al. (1956), e mais tarde por Ball e

Hall (1967). O objetivo do algoritmo k-means é dividir um objeto em n dimensões em k partições (ou clusters) de forma a minimizar a soma dos quadrados das diferenças dos itens pertencentes a cada cluster em relação ao centro. O k-means forma uma estrutura plana, onde a semelhança entre um par de objetos é definida pela sua distância, essa distância é definida pela distância euclidiana.

2.2.3 Rede Neural Convolucional

A Rede Neural Convolucional (CNN) é uma arquitetura de aprendizado profundo que foi inspirada no mecanismo natural de percepção visual das criaturas vivas. Em 1959, os cientistas David Hubel e Torsten Wiesel (“[Receptive fields and functional architecture of monkey striate cortex](#)” 1968) fizeram uma descoberta significativa ao identificar células no córtex visual de animais que são responsáveis por detectar luz em campos receptivos. Esta descoberta foi um marco na compreensão da visão humana e inspirou várias pesquisas subsequentes, incluindo a proposta do neocognitron em 1980 por Kunihiko Fukushima(Fukushima 1980). Essa arquitetura é amplamente utilizada em tarefas de visão computacional e classificação de imagens, onde a capacidade de processar e aprender padrões em dados de imagem é crucial.

As Redes Neurais Convolucionais (CNNs) processam dados em diferentes formatos, como imagens, usando conexões locais, compartilhamento de pesos, pooling e múltiplas camadas não lineares. Elas são adaptáveis a várias dimensões de entrada, como 1D para sinais, 2D para imagens e 3D para vídeos ou dados volumétricos (Goodfellow, Bengio e Courville 2016 e W. Ma et al. 2019). CNNs 1D exploram informações em uma dimensão, investigando correlações espectrais em pontos específicos do pixel. Já as CNNs 2D analisam características espaciais nas imagens, enquanto as CNNs 3D combinam características espaciais e espectrais. Utilizando convoluções 3D, um filtro tridimensional é aplicado aos dados, movendo-se pela largura (w), altura (h) e profundidade (d) para identificar possíveis padrões (W. Ma et al. 2019). Neste trabalho, usamos duas dimensões espaciais e uma temporal; portanto, foi implementada uma arquitetura usando redes convolucionais 3D.

Nas redes convolucionais, o reconhecimento de características segue uma abordagem supervisionada, como descrito por LeCun, Bengio et al. (1995). O processo de agrupamento (Pooling) é essencial nas redes neurais convolucionais (CNNs), reduzindo a dimensionalidade das matrizes. Esse método mapeia cada partição para um único valor (Cheung et al. 2020). Diversas formas de agrupamento existem, como o pooling máximo que seleciona os valores mais altos de cada janela, o pooling mínimo que considera os valores mais baixos e o pooling médio que considera o valor médio de cada janela. A extração de características em redes convolucionais é realizada por meio de neurônios que recebem

entradas sinápticas e detectam características locais. Cada camada da rede é composta por vários mapas de características, onde cada mapa é um plano no qual os neurônios compartilham um conjunto de pesos sinápticos comuns. Esse compartilhamento de pesos é acompanhado por funções de ativação, que introduzem não linearidade nas saídas das camadas, permitindo que a rede aprenda e represente relações complexas nos dados. As principais funções de ativações são:

A função linear multiplica um peso e gera valores em todo o domínio dos números reais. No entanto, sua principal limitação é a incapacidade de capturar relações complexas entre os dados, devido à sua derivada parcial constante, o que limita o espaço de funções que podem ser representadas apenas por transformações lineares. A função de ativação linear pode ser aplicada na camada de saída em problemas de regressão, permitindo o mapeamento da saída para qualquer valor real.

A função ReLU (do inglês Rectified Linear Unit) é uma função de ativação bastante usada em redes neurais, ela retorna valores que variam de 0 ao infinito. A função ReLU retorna zero (0) para os valores negativos, o que tende a apagar alguns neurônios durante o processo de forward, aumentando assim a velocidade do treinamento. Esta função introduz uma relação não-linear à rede, possibilitando que as redes neurais aprendam representações mais complexas.

A função sigmoide gera valores entre 0 e 1, introduzindo não linearidade nas camadas de uma rede neural. Sua principal vantagem reside na sua derivada, que é máxima próximo do zero. Essa característica é particularmente útil para calcular probabilidades em problemas de classificação binária.

A função Softmax é uma extensão da função sigmoidal, sendo aplicada principalmente nas camadas de saída de redes neurais para problemas de classificação não-binária. Ela transforma os logits em probabilidades, garantindo que a soma das probabilidades seja igual a 1. No entanto, a função Softmax não é comumente usada em camadas intermediárias, pois produz valores apenas no intervalo de 0 a 1.

2.3 Estado da Arte do Uso dos Modelos Supervisionados na Previsão das Tempestades Severas

Collins e Tissot (2007) desenvolveram uma rede neural artificial para incorporar dados de modelos numéricos para aprimorar a previsão das tempestades. Os seus resultados indicaram a capacidade de previsão da rede neural artificial. Alves et al. (2017) criaram uma rede neural artificial (ANN) treinada com o algoritmo de retro-propagação Levenberg-

Marquardt para classificar a modelagem relacionada à previsão das descargas elétricas na região amazônica. Os seus resultados mostraram que a ANN foi capaz de identificar corretamente a classe à qual um evento pertencia em relação as categorias de ocorrência e ausência das descargas elétricas. Geng et al. (2019) sugeriram um modelo orientado por dados baseado em redes neurais chamado LightNet para a previsão das descargas elétricas. Para os autores, o modelo LightNet extrai características espacialmente e temporalmente por meio de codificadores duplos, e essas características são alimentadas em um codificador para fazer previsões. Jahangir, Reineh e Abolghasemi (2019) baseados nos dados de estações hidrométricas e pluviométricas e imagens de satélite desenvolveram uma rede neural para previsão das inundações na Bacia do Rio Kan, no Irã. Da mesma maneira, em Yeditha et al. (2020) desenvolveram um método de previsão utilizando produtos de precipitação por satélite e modelos de aprendizado de máquina baseados em ondulas aplicados para prever precipitações extremas. A abordagem proposta foi testada na previsão de inundações na bacia do rio Vamsadhara, na Índia. Os mesmos autores (Yeditha et al. 2020) documentaram que a previsão em tempo real de eventos extremos de inundação com prazos mais longos é difícil principalmente em regiões com poucos pluviômetros, e, nessas situações, a precipitação por satélite aplicada a modelos de inteligência artificial pode ser uma alternativa melhor. Nyap, Waheeb e Luokse (2020) desenvolveram um algoritmo para previsão das descargas elétricas usando a técnica de lógica difusa. Os testes realizados na sua pesquisa mostraram uma boa capacidade do seu algoritmo na previsão das descargas elétricas com 1 hora de antecedência, com uma precisão de 95%. Makgatho, Ajoodha e Hunt (2020) utilizaram os dados meteorológicos como: umidade, temperatura e pressão atmosférica como dados de entrada dos seus dois modelos (regressão logística e redes neurais) para a previsão de ocorrência das descargas elétricas. Os seus resultados indicaram 95% de precisão para redes neurais e 87% para regressão logística. Leinonen, Hamann e Germann (2022) com objetivo de prever a ocorrência das descargas elétricas, criaram uma arquitetura recorrente-convolucional que permite reconhecer e prever o desenvolvimento da convecção no tempo e no espaço, incluindo o movimento de células de tempestades, usados vários dados de entrada: radar, satélite, detecção de raios, previsão numérica do tempo e um modelo de elevação digital. Chen et al. (2022) desenvolveram um modelo de rede neural convolucional LSTM de dois fluxos para a previsão imediata de precipitação. Os seus resultados demonstram um desempenho eficaz do modelo em um conjunto de dados de grande escala.

Schmidt (2023) desenvolveram um modelo de previsão de curto prazo que utiliza redes neurais convolucionais (CNNs) para produzir previsões de granizo severo para a próxima hora.

Dados e Metodologia

3.1 Dados

GLM (do inglês, Geostationary Lightning Mapper) é um sensor que está a bordo do satélite GOES-16 (GOES do inglês, Geostationary Operational Environmental Satellite), a uma altura de 36000 km. GLM é um detector óptico de transientes (OTD), que mede continuamente a atividade total dos raios nas Américas. Goodman et al. (2013) traz suas características. O GLM usa um dispositivo de imagem acoplado a carga que opera no infravermelho, no comprimento de onda de aproximadamente 777,4 nm. Suas imagens são geradas a cada 2 ms em uma matriz de 1372 por 1300 pixels. Além disso, o GLM mapeia toda a extensão da nuvem iluminada por relâmpagos individuais, com uma resolução de 8 por 8 km no nadir, variando de 13 ou 14 km. As atualizações temporais fornecidas pelo GLM permitem mapear a estrutura do flash, com uma eficiência de detecção de pelo menos 70 por cento.

Os dados de reanálise do Era-5, a quinta geração do Centro Europeu de Previsões Meteorológicas de Médio Prazo (ECMWF), foram lançados pelo Copernicus Climate Change Service. Era-5 apresenta conjunto de dados meteorológicos obtidos a partir de modelos de circulação global da atmosfera, radiossonda, aviões e satélites. Os dados de Era-5 são disponibilizados através do Climate Data Store, e são atualizados regularmente. Eles são organizados em grades regulares de latitude e longitude, com uma resolução espacial de 0,25 graus por 0,25 graus e uma resolução temporal de 1 hora. Esses dados abrangem diversos parâmetros atmosféricos distribuídos em 37 níveis de pressão. Era-5 foi lançada pelo ECMWF (2017) em substituição da reanálise Era-Interim. Neste trabalho foram usados os dados de Era-5 para avaliar o perfil vertical da atmosfera em ambientes de ocorrência (ou não ocorrência) das tempestades elétricas. Os parâmetros atmosféricos usados neste trabalho incluem as informações de temperatura, pressão atmosférica, umidade do ar e os ventos.

3.1.1 Descrição dos Parâmetros Atmosféricos Usados na Pesquisa

Os parâmetros atmosféricos são medidas que descrevem diferentes aspectos do estado da atmosfera. Eles fornecem informações essenciais para entender a dinâmica e a termodinâmica das condições atmosféricas em locais e momentos específicos. A ocorrência das descargas elétricas é influenciada por variáveis dinâmicas e termodinâmicas da atmosfera (Stolz, Rutledge e Pierce 2015 e Igel e Heever 2015). Moller (2001) descreveu os ingredientes para a formação de convecção profunda, associado com a formação das tempestades elétricas, que incluem a presença de instabilidade condicional, existência de moderado a intenso cisalhamento vertical do vento, oferta de umidade na baixa troposfera e um mecanismo de levantamento das parcelas do ar.

CAPE (do inglês, Convective Available Potential Energy) é a energia potencial que a parcela de ar tem no momento que é elevada até o nível de convecção espontânea, e se torna mais quente comparativamente com o ar do ambiente, experimentando a flutuabilidade positiva (Riemann-Campe, Fraedrich e Lunkeit 2009). CAPE é um parâmetro atmosférico bastante usado nos estudos das tempestades severas (exemplo: Taszarek et al. 2019), pelo fato dos seus valores estarem associados aos movimentos relativos dos hidrometeoros dentro da nuvem dependem da velocidade vertical da corrente ascendente (E. N. Rasmussen e Blanchard 1998).

O índice KI é usado para avaliar o potencial convectivo, que é baseado na taxa vertical da temperatura, conteúdo e extensão vertical da umidade em baixos níveis (850 e 700 hPa) (George 1960). O KI também analisa taxa vertical da temperatura nos níveis de 850 e 500 hPa e o teor de umidade em 700 hPa.

O índice TT é usado para avaliar a profundidade de umidade na camada limite da atmosfera (Miller 1975). De acordo com Henry (1987), o índice TT envolve a taxa de variação vertical de temperatura entre os níveis de 850 a 500 hPa, e é influenciado por sistemas sinóticos que fazem com que os seus valores aumente ou diminua. Recentemente Mandanda (2021) documentaram que a instabilidade atmosférica é indispensável na formação das tempestades elétricas, onde notaram que o aumento os valores de CAPE, TT e KI associa-se com o aumento das descargas elétricas.

DLS (do inglês, Deep Layer Shear) traduzido como cisalhamento profundo, definido como o modulo da diferença vetorial entre o vento a 6000 m de altura e o vento médio nos primeiros 500 m acima do solo (E. N. Rasmussen e Blanchard 1998), é um parâmetros bastante usado na literatura para avaliar o potencial de tempo severo. Além de DLS, também existe o cisalhamento vertical do vento em baixos níveis (LLS, do inglês Low Level Shear) que é bastante documentado na literatura (por exemplo: Groenemeijer e Van Delden 2007). LLS é definido como o modulo da diferença vetorial entre o vento a 1000 m e o vento da superfície. LLS e DLS são parâmetros bastante usados para prever

condições meteorológicas severas. Craven, Brooks, Hart et al. (2004) constataram que altos valores de LLS são frequentemente associados a tornados e a ocorrência de granizos.

Velocidade vertical em 500 hPa expressa se a grande escala é propícia ou não a movimentos verticais do ar em 500 hPa. Os movimentos ascendentes associam-se a instabilidade atmosférica e a subsidência a estabilidade atmosférica (Uccellini e Johnson 1979).

A divergência de umidade em 850 hPa determina o fluxo de umidade em baixos níveis (850 hPa) (Banacos e D. M. Schultz 2005). Valores positivos representam a divergência de umidade e valores negativos, a convergência de umidade em 850 hPa.

3.2 Metodologia

Para alcançar o objetivo desta pesquisa, foi necessário analisar primeiro a distribuição das descargas elétricas em cada ponto de grade na América do Sul. Com base nos bancos de dados das médias mensais e ciclos diurnos das descargas elétricas no período entre 2018 e 2020, aplicou-se para cada banco de dados o algoritmo k-means (Fotouhi e Montazeri-Gh 2013). A primeira análise de cluster resultou em duas figuras que apresentam regiões com diferentes ciclos anuais e diurnos das descargas elétricas na América do Sul.

Em seguida, juntou-se os dados do ciclo diurno e as médias mensais das descargas elétricas formando um conjunto de dados. Posteriormente foi aplicada a técnica de redução de dimensionalidade (PCA), que permitiu reduzir o conjunto de dados e identificar os componentes mais significativos que explicam a maior parte dos dados. Após a aplicação da técnica PCA, foram selecionados os primeiros 5 componentes significativos. Em seguida, foi aplicada novamente a análise de agrupamento. A última análise de agrupamento permitiu dividir a América do Sul em várias sub-regiões. Para alcançar o objetivo geral do trabalho, optamos por focar apenas na região da Bacia do Prata para realizar a previsão das descargas elétricas.

Para a construção do modelo de previsão das descargas elétricas na Bacia do Prata, foram utilizados os dados horários das descargas elétricas como saída da rede convolucional 3D. Já os dados de entrada foram baseados em parâmetros atmosféricos descritos anteriormente. Pelo fato da Bacia do Prata abranger uma grande área, ao invés de usar o conjunto de dados horários de 2018 a 2020, o intervalo foi reduzido para seis (6) meses (junho a dezembro de 2020), com o objetivo de tornar a base de dados mais manejável e possível de ser analisada em uma instância de GPU da plataforma Google Colab¹. A separação dos dados foi feita de seguinte forma: 70% para treinamento, 15% para validação e 15% para teste. A arquitetura da rede desenvolvida é constituída por camadas convolu-

¹<https://colab.research.google.com/>

cionais 3D e camadas LSTM (do inglês Long Short-Term Memory). A camada LSTM é um tipo de camada em redes neurais recorrentes (RNNs) que é especialmente eficaz para lidar com sequências de dados. Nas camadas ocultas da rede usou-se a função de ativação ReLU. Os dados das descargas elétricas são desbalanceados pelo fato de a grande maioria dos horários não possuírem atividades elétricas, por causa disso, aplicou-se uma técnica de balanceamento dos dados que consistiu em atribuir maior peso no treinamento aos valores das descargas elétricas superiores a zero. Essa atribuição foi feita na função de perda.

Para a avaliação do modelo desenvolvido foram analisadas as seguintes métricas estatísticas:

O Erro Médio Absoluto (MAE, do inglês Mean Absolute Error) é uma métrica usada para avaliar a precisão de um modelo de previsão. Ele é calculado como a média das diferenças absolutas entre os valores reais e previstos por um modelo. O MAE é resistente a outliers, o que significa que ele não é fortemente afetado por valores extremos nos dados. O MAE é dado por seguinte equação:

$$\text{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.1)$$

Onde y_i é o valor real da i -ésima observação e \hat{y}_i é o valor previsto para a i -ésima observação.

A Raiz do Erro Quadrático Médio (RMSE, do inglês Root Mean Square Error) é uma métrica que avalia a precisão do modelo de previsão, e é sensível a magnitude dos erros. Um RMSE menor indica uma maior precisão na previsão. A equação do RMSE é dada por:

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{N}} \quad (3.2)$$

Resultados e Discussão

4.1 Análise da Distribuição das Descargas Elétricas na América do Sul

A Figura 4.1 ilustra a aplicação do método do cotovelo para determinar o número ótimo de clusters na análise da distribuição dos raios na América do Sul. Este método avalia a variância dos dados em relação ao número de clusters, identificando o ponto de inflexão como o número ideal de clusters. Na mesma figura observa-se que de 0 a 10 há uma variação acentuada nos dados, enquanto após 10 a variação não é muito significativa. Com base nos objetivos da pesquisa decidiu-se usar 4 clusters, por ser um bom equilíbrio entre a redução do número de clusters e a representatividade de cada um deles (Figuras 4.2 e 4.4).

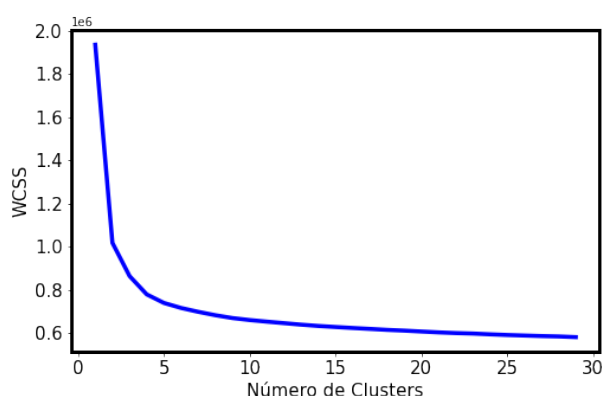


Figura 4.1: Método de cotovelo. Onde WCSS é a distância total dos pontos dos dados em relação aos seus centroides.

4.1.1 Análise de Cluster Baseado nas Médias Mensais das Descargas Elétricas na América do Sul

A Figura 4.2 apresenta as áreas na América do Sul com uma distribuição uniforme do ciclo anual de descargas elétricas, identificadas por cores distintas. Destacam-se quatro regiões com padrões anuais diferentes de descargas elétricas. Cluster 1 (vermelho) inclui a Argentina, o Paraguai, a parte oeste da Bolívia, uma parte do Peru e no Brasil, sobretudo nos estados de Mato Grosso, Tocantins, Minas Gerais, parte oeste da Bahia, região litoral de São Paulo e Rio de Janeiro. O Cluster 2 (azul) abrange principalmente a Amazônia, uma parte do Sul do Brasil e o Uruguai. Já o Cluster 3 (verde) engloba a maior parte do Nordeste Brasileiro, Sul da Argentina, Chile, Peru, Equador, parte leste da Bolívia, Suriname e Guiana. Por fim, o Cluster 4 (roxo) inclui a porção norte da Colômbia e Venezuela, especialmente na região do lago Maracaibo.

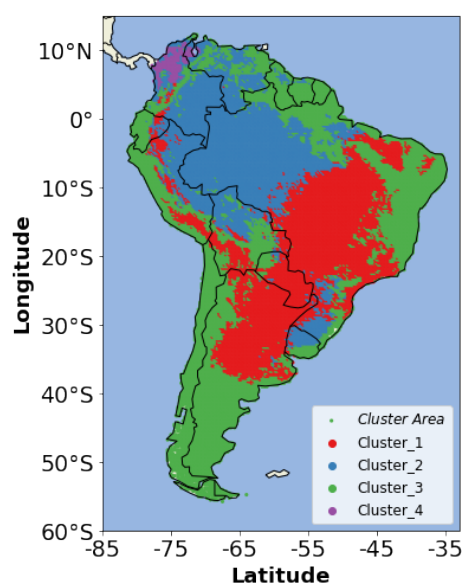


Figura 4.2: Análise de cluster na América do Sul baseando-se no ciclo anual. Onde cluster 1 (vermelho), cluster 2 (azul), cluster 3 (verde), cluster 4 (roxo).

A região do cluster 1 (Figura 4.3a) apresenta maior atividades elétricas na primavera e no verão. Na estação chuvosa há maior disponibilidade de calor e umidade na atmosfera. No entanto, no inverno espera-se baixa atividade elétrica pelo fato de que nessa estação a atmosfera tende a ser mais estável relativamente.

De acordo com Peixoto e Oort (1992), o Paraguai e a parte leste da Argentina são afetados por fortes gradientes meridionais importantes na formação de sistemas frontais. A região próxima aos Andes, na Argentina, é um local de ciclogênese (Satyamurty e De Mattos 1989). A ciclogênese está associada à convergência do ar em baixos níveis

da atmosfera, ambiente em que podem se desenvolver as tempestades elétricas severas. K. L. Rasmussen, Zuluaga e Houze Jr (2014) documentaram que na primavera tem um máximo de raios no Nordeste da Argentina e Paraguai associado à convecção profunda. No decorrer do tempo, a atividade elétrica tende a se mover para região sudoeste da Argentina, próximo dos Andes, enquanto os sistemas convectivos de mesoescala se desenvolvem e se movem para leste da Argentina. Na região Sul da América do Sul verifica-se uma contribuição adicional da ciclogênese cruzando a Cordilheira dos Andes (K. L. Rasmussen e Houze 2011).

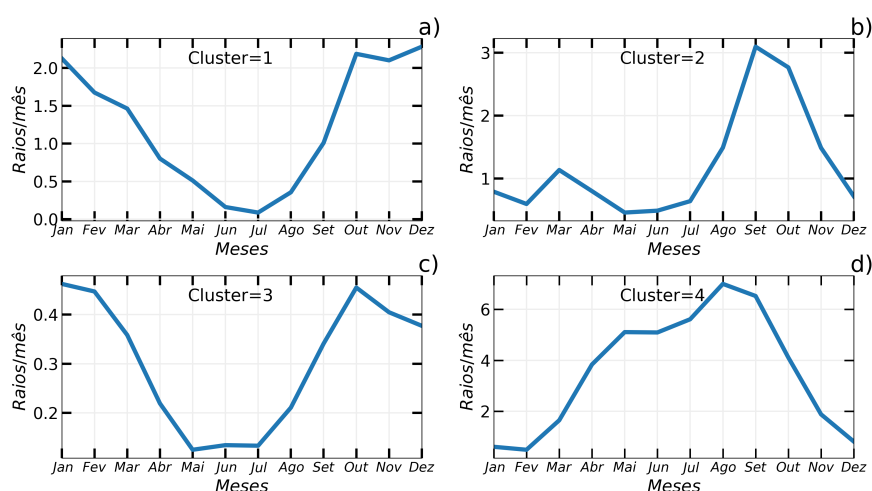


Figura 4.3: Ciclo anual das descargas elétricas em cada região(Clusters).

No cluster 2, que engloba maior parte da Amazônia, há uma maior atividade elétrica na primavera (Figura 4.3b). Os resultados semelhantes foram documentados por Diaz, Ortiz e Roman (2022), Oda et al. (2022) e Morales Rodriguez (2019), que observaram nos seus resultados a máxima atividade elétrica na primavera, principalmente na bacia amazônica. Segundo Morales Rodriguez (2019), na estação chuvosa, a convecção na região Amazônica tende a ser mais generalizada e inibe o desenvolvimento das tempestades elétricas. Portanto, a atividade elétrica na Amazônia é influenciada pelas transições entre as estações seca e chuvosa (Saraiva et al. 2016).

No verão, apesar de ocorrer baixa atividade elétrica na Amazônia, uma parte dos raios são influenciados pela entrada das linhas de instabilidade (Uvo e C. Nobre 1989). Para Carlos A Nobre, Sellers e Shukla (1991) e Da Rocha et al. (2009), a convecção na região Noroeste da Amazônia está associada à condensação de ar úmido transportado pelos ventos alísios, que sofrem levantamento orográfico perto dos Andes. De acordo com Garreaud e Wallace (1997) e Da Rocha et al. (2009) a convecção próximo dos Andes é observada durante a noite, como resultado da convergência do escoamento da montanha com os alísios de nordeste sobre a região da bacia amazônica. Na região de cluster 3 a

maior atividade elétrica ocorre na primavera e no verão, e baixo no inverno (Figura 4.3c). Nessa região, as descargas elétricas no verão ocorrem possivelmente devido ao intenso aquecimento radiativo, que favorece o desenvolvimento da convecção, principalmente no Nordeste Brasileiro. Nas regiões litorais-Oceânicas como Sul e Sudeste Brasileiro, as atividades elétricas estão associadas com ciclones extratropicais de escala sinótica e as frentes frias (Albrecht et al. 2016).

Oda et al. (2022) também observaram na sua pesquisa a ocorrência de maior atividade elétrica no verão. Para Virts et al. (2013) a atividade elétrica no verão é influenciada principalmente pelo aquecimento diurno na América do Sul. Um dos sistemas que influenciam na ocorrência da convecção na região do Chile e uma parte da região Oeste da Argentina é o posicionamento do Anticiclone Sul do Pacífico Sul (ASPS), que origina os movimentos subsidentes (Reboita et al. 2010). Para os mesmos autores, no inverno, os ventos de oeste ao Sul do ASPS atingem a cordilheira dos Andes e são forçados a ascender. Portanto, nesse processo o escoamento é resfriado adiabaticamente, favorecendo assim a formação de nuvens de tempestades elétricas.

O cluster 4 engloba o Lago de Maracaibo e parte Norte da Colômbia. A maior distribuição dos raios nesta região é observada no verão boreal (Figura 4.3d). Resultados semelhantes foram documentados por Albrecht et al. (2016), na qual observaram que a maior atividade elétrica no Lago Maracaibo ocorre do verão ao outono boreal. A atividade elétrica na Colômbia é influenciada principalmente pela topografia, padrões de vento e disponibilidade de umidade que favorecem a ocorrência da convecção profunda (Bürgesser, Nicora e Avila 2012). Analisando a magnitude das médias mensais das descargas elétricas, pode-se notar que de todas regiões (clusters) analisadas na América do Sul, a região de Cluster 4 (Figura 4.3d) apresenta altos valores. Esses resultados estão de acordo com Albrecht et al. (2016) e Peterson (2021), que apresentaram nos seus resultados as regiões que apresentam maior taxa de raios no mundo, na qual o Lago Maracaibo é considerado como sendo o local onde ocorre maior taxa de raios na terra.

4.1.2 Análise de Cluster Baseado no Ciclo diurno das Descargas Elétricas na América do Sul

A Figura 4.4 mostra as regiões com ciclo diurno homogêneo dos raios na América do Sul. Através dela podem se destacar 4 regiões com ciclo diurno diferente. Dentre elas, o cluster 1 (cor vermelha), que engloba Chile, Peru, parte Oeste e Sul da Argentina, parte central de Venezuela, Equador, Suriname, Guiana, Nordeste Brasileiro, estado de Pará, pequena área de Minas Gerais, parte litoral do Rio Grande do Sul e Uruguai; o cluster 2 (azul), que inclui principalmente a maior parte da bacia Amazônica, Mato Grosso, Mato Grosso do Sul, Tocantins, Goiás, litoral de São Paulo e Rio de Janeiro; o cluster 3 (verde),

que engloba pequena parte do norte da Colômbia e o Lago de Maracaibo; e o cluster 4 (roxo), que inclui principalmente a parte leste da Argentina, Paraguai, Uruguai e parte Oeste do Rio Grande do Sul.

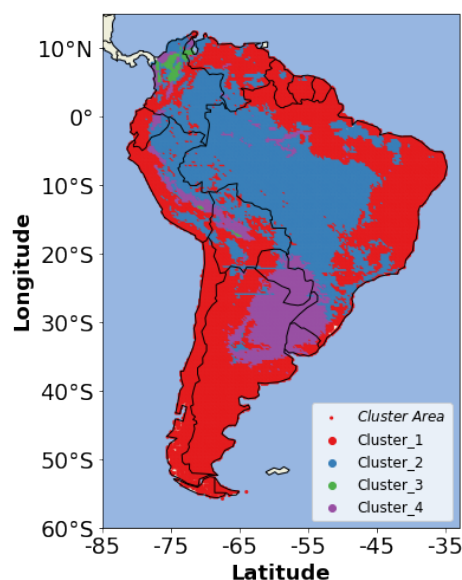


Figura 4.4: Análise de cluster na América do Sul (baseando-se na variação diurna das descargas elétricas). Onde cluster 1 (vermelho), cluster 2 (azul), cluster 3 (verde), cluster 4 (roxo).

A Figura 4.5a mostra o ciclo diurno das descargas elétricas na região de cluster 1, na qual observa-se maior pico das atividades elétricas no período da tarde (entre 12 a 18 horas locais). Esses resultados apresentam uma concordância com Oda et al. (2022), que conclui na sua pesquisa que as regiões como Sudeste Brasileiro e o estado da Bahia apresentam o máximo pico das atividades elétricas no período entre 12 e 18 horas.

A maior atividade elétrica nas áreas de Cluster 2 ocorre no período da tarde. A fonte de convecção na bacia Amazônica deve-se à combinação da convergência dos ventos de leste na superfície (Romatschke e Houze 2010) combinada com os efeitos do aquecimento diurno. DiGangi et al. (2022) encontrou máxima atividade elétrica no período da tarde na região da bacia amazônica e na parte central do Brasil. A atividade elétrica na Colômbia ocorre entre o meio-dia e de manhã cedo nas regiões do Caribe, Andinha e Pacífico (DiGangi et al. 2022).

Na região de cluster 3 (Figura 4.5c) verifica-se a máxima atividade elétrica no período noturno. Garreaud e Wallace (1997) e Virts et al. (2013) documentaram que a convecção profunda observada nos vales ao norte da Cordilheira dos Andes, incluindo o lago Maracaibo, é influenciada pelas circulações térmicas locais (brisas terrestres-marítimas e brisas montanhosas-vale), que ocasionam as convergências em baixos níveis. Essas circulações

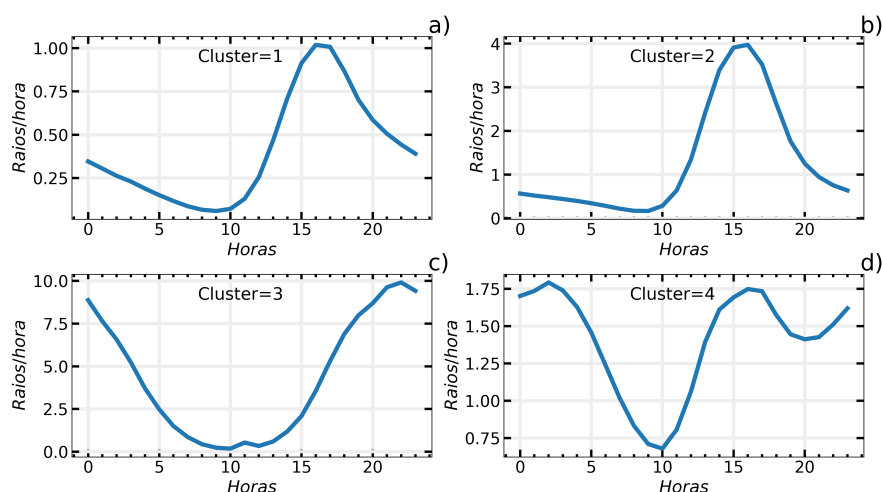


Figura 4.5: Ciclo diurno (horário local) das descargas elétricas em cada região (clusters).

térmicas locais favorecem a convecção no período da manhã na região de Colômbia e Lago Maracaibo (Cecil, D. E. Buechler e Blakeslee 2014).

Diaz, Ortiz e Roman (2022) concluíram na sua pesquisa que as atividades elétricas na Colômbia apresentam dois picos dependendo da região, por exemplo nas regiões do Caribe, Andina e Pacífico as atividades elétricas ocorrem entre meio-dia e de manhã cedo. A maior atividade elétrica na Colômbia ocorre na região do Caribe, localizado ao longo dos Andes e bacias hidrográficas dos principais Rios colombianos.

Na região de cluster 4 (Figura 4.5d) nota-se dois períodos de máxima atividade diurna das descargas elétricas, o maior pico, observado às 2 horas locais, está associado principalmente pela atuação de sistemas convectivos de mesoescala. Esses SCM apresentam maior atividade no período noturno. A parte Leste da Argentina, Paraguai e Uruguai são regiões propensas às atividades noturnas dos Complexos Convectivos de Mesoescala. Durkee, Mote e Shepherd (2009) mostram nos seus resultados a frequência de ocorrência dos Complexos Convectivos de Mesoescala (CCM) na América do Sul, onde verifica-se maior frequência de CCM na região de cluster 4 (Figura 4.5d), principalmente no Norte da Argentina, Paraguai e Sul do Brasil. Os CCMs nas regiões extratropicais da América do Sul são formados sobretudo com a presença de Jatos de Baixos níveis que possibilitam forte advecção de ar quente e úmido procedente da bacia amazônica (Velasco e Fritsch 1987; Marengo et al. 2004, Salio, Nicolini e Zipser 2007).

Morales Rodriguez (2019) e Albrecht et al. (2016) documentaram que a região de cluster 4 apresenta dois picos de máxima atividade diurna das atividades elétricas, dos quais um máximo noturno associado aos CCMs e outro pico durante o período da tarde associado a maior insolação. Para Geerts et al. (2017) os SCM podem ocorrer várias vezes por semana e podem persistir por até 12 horas, produzindo raios ao longo da sua vida.

De acordo com Oda et al. (2022), as atividades elétricas na região Sul do Brasil atingem o seu pico máximo nas primeiras horas, associadas provavelmente pela atuação noturna dos SCM.

4.1.3 Análise de Componentes Principais

A análise de componentes principais (PCA) foi aplicada para um banco de dados que incluiu as médias mensais e as médias horárias das descargas elétricas na América Sul. Ou seja, foram juntadas as séries temporais usadas nas Figuras 4.3 e 4.5. Portanto, depois de juntar os dados, foi formado um banco de dados em que foi aplicada uma análise de componentes principais para reduzir sua dimensionalidade às cinco principais componentes, que separam melhor as regiões com características semelhantes do ciclo anual e diurno (Figura 4.6).

Através da Figura 4.6, pode se notar claramente regiões influenciadas por sistemas meteorológicos diferentes. Cluster 1 no geral engloba toda região Amazônica, as atividades elétricas nessa região são influenciadas pela dinâmica e termodinâmica local. Cluster 2 engloba Argentina, Paraguai, Parte Oeste de Uruguai e parte Oeste do Rio Grande do Sul, nessas regiões as atividades elétricas são influenciadas principalmente pelos Complexos Convectivos de Mesoescala (Durkee, Mote e Shepherd 2009). De uma forma geral a análise de cluster dos primeiros componentes apresenta melhor separação. Com base nessas separações apresentadas na Figura 4.6, o próximo passo consiste em analisar os parâmetros atmosféricos e criar uma rede neural para prever as atividades elétricas na região do cluster 2.

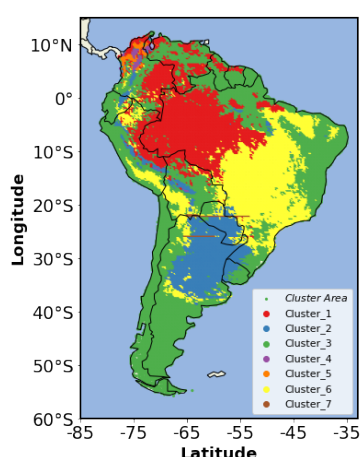


Figura 4.6: Análise de cluster na América do Sul, baseado nos 5 componentes principais.

4.2 Previsão das Descargas Elétricas na Bacia do Prata

A previsão das descargas elétricas foi feita somente na região de cluster 2 (Figura 4.6) que é a região da Bacia do Prata, que engloba Argentina, Paraguai, Uruguai e o Sul do Brasil. A previsão fez-se usando os dados horários das descargas elétricas (dados de saída) e dados de reanálise do Era-5 como: CAPE, índice TT, KI, umidade relativa em 850 hPa, cisalhamento vertical do vento próximo à superfície (entre ventos a 10 m e ventos a 975 hPa), pressão a nível médio do mar (MSLP), divergência dos ventos em 850 hPa, velocidade vertical em 500 hPa e cisalhamento dos ventos próximo à superfície até 450 hPa (DLS). Os dados de reanálise do Era-5 serviram como dados de entrada do modelo.

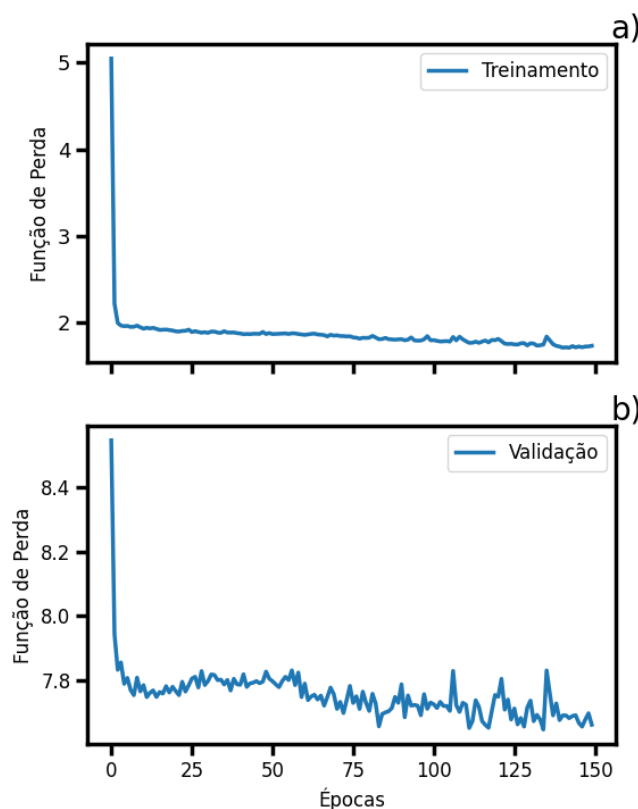


Figura 4.7: Função de perda do (a) treinamento e (b) da validação.

A Figura 4.7 mostra o progresso da função de perda em relação às épocas durante o período de treinamento e validação da rede convolucional 3D. Os resultados mostram que os valores da função de perda diminuem na medida que aumentam as épocas, porém essa diminuição é lenta principalmente depois de atingir valor 2 durante o treinamento (Figura 4.7a) e na validação (Figura 4.7b). Durante a validação a magnitude do erro tende a ser maior (Figura 4.7a) comparado com a função de perda durante o treinamento (Figura

4.7b). Uma das possíveis causas das diferentes magnitudes das funções de perdas durante o treinamento e validação é o desbalanceamento dos dados das descargas elétricas, que apresentam maior número de dados nulos.

A Figura 4.8 mostra o acumulado das descargas elétricas do mês de setembro do ano 2020. Através dela, verifica-se a região com maior densidade das descargas elétricas (-60 até -55). Esses dados correspondem a uma parte dos dados usados no treinamento do modelo convolucional 3D.

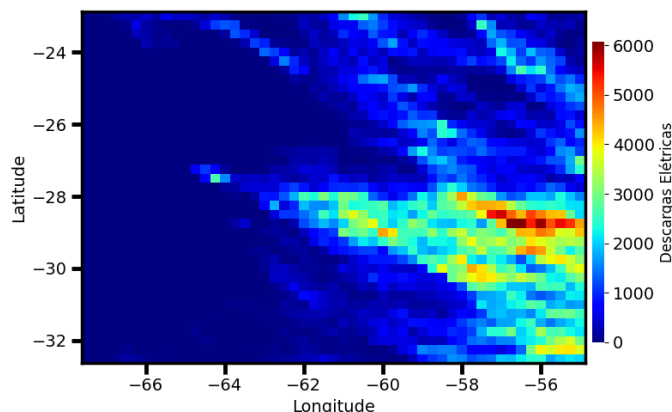


Figura 4.8: Acumulado das descargas elétricas do mês de Setembro do ano 2020.

A Figura 4.9 mostra os resultados obtidos durante o teste do modelo desenvolvido. Através da figura 4.9 podemos notar que tanto nos dados observados (Figura 4.9.b) quanto nos previstos (4.9a) apresentam região similar de maior densidade das descargas elétricas, sobretudo entre latitude -62 até -55 graus.

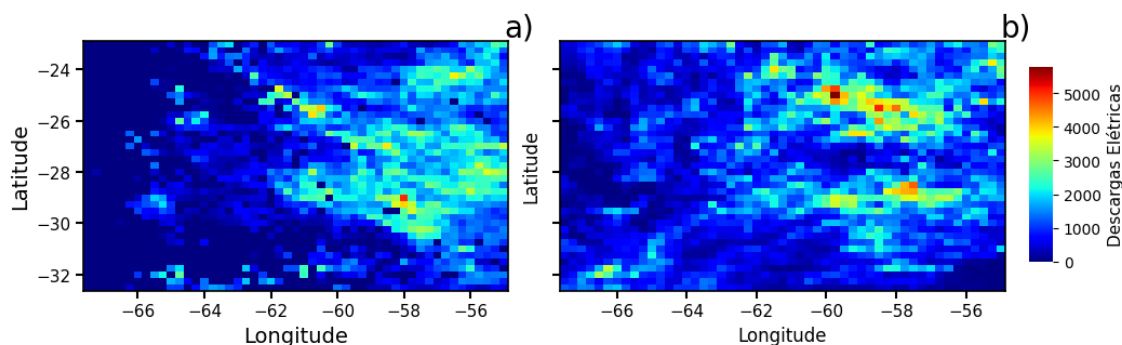


Figura 4.9: Número total das descargas elétricas em cada ponto de grade. Onde (a) descargas elétricas previstas e (b) observadas.

A Figura 4.10 mostra a serie temporal das descargas elétricas previstas (linha alaranjada) e observadas (linha azul). Portanto, através desta Figura 4.10, percebe-se que, mesmo não acertando o horário com exatidão, o modelo se aproxima consideravelmente dos momentos em que há descargas elétricas. Uma observação semelhante pode ser feita

na Figura 4.9, em que o modelo acerta a região com o maior número de raios, mas, não necessariamente, o ponto da malha que os receberá.

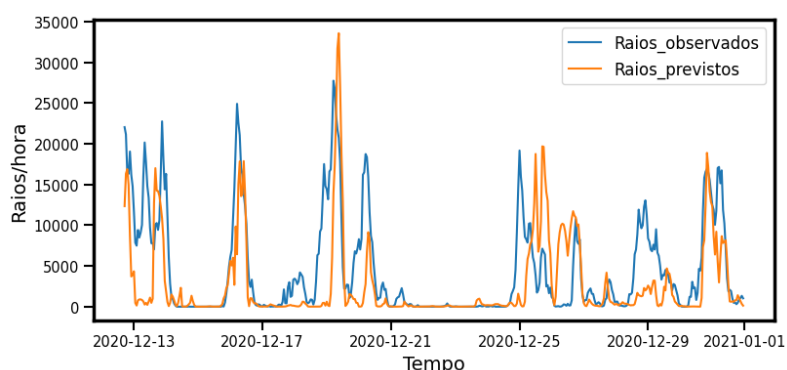


Figura 4.10: Acumulado horário das descargas elétricas. Onde a linha laranja são descargas elétricas previstas e a linha azul são as descargas elétricas observadas.

A Figura 4.11 Apresenta as métricas estatísticas que avaliam o desempenho do modelo desenvolvido. O Erro Absoluto Média (MAE) é de 2.40 raios por hora, um valor relativamente baixo, indicando um bom desempenho do modelo (Figura 4.11). No entanto, a Raiz Quadrada do Erro Médio Quadrado (RMSE) é de 22.47 raios por hora, uma métrica que é fortemente influenciada pelos valores extremos, neste caso, representados por dias com um alto número de descargas. O modelo demonstra maior acerto nos horários sem descargas elétricas (Figura 4.10), o que pode ser uma das razões para o baixo valor do MAE. Contudo, para os dias com descargas elétricas, a precisão do modelo diminui devido à falha em capturar a magnitude de alguns picos (Figura 4.10), o que impactou no valor de RMSE.

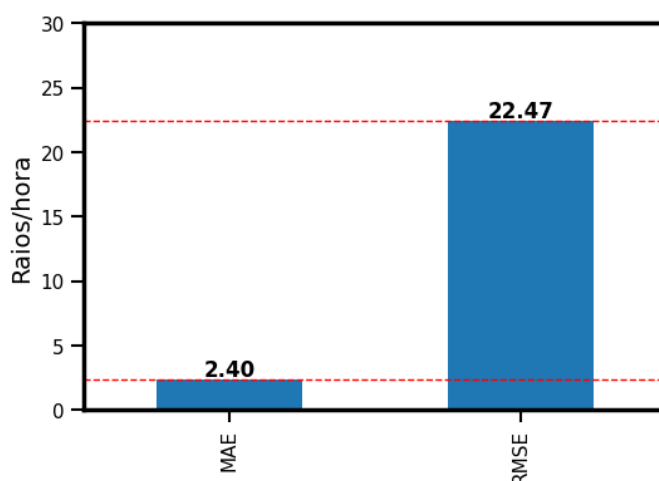


Figura 4.11: Métricas estatísticas

Conclusão

Este trabalho analisou as distribuições das descargas elétricas na América do Sul, que foi crucial na delimitação da área de estudo. O objetivo principal do trabalho foi de prever a ocorrência das descargas elétricas na Bacia do Prata usando redes convolucionais 3D. Com base nos resultados obtidos pode se concluir o seguinte:

- A análise de cluster conseguiu separar as regiões com o mesmo ciclo diurno e anual das descargas elétricas na América do Sul.
- A análise de cluster dos primeiros cinco componentes conseguiu separar as regiões que apresentam mesma climatologia dos sistemas meteorológicos, principalmente a bacia do Prata que engloba os seguintes países: Argentina, Paraguai, Sul do Brasil e Uruguai.
- O modelo demonstrou um desempenho eficaz na identificação das áreas de maior concentração de descargas elétricas.

5.1 Limitações e Trabalhos Futuros

A principal limitação identificada neste estudo é a escassez de recursos computacionais que são capazes de lidar com o processamento de grandes volumes de dados espaciais utilizando redes neurais convolucionais 3D. Embora este trabalho esteja de acordo com várias pesquisas, é importante ressaltar algumas sugestões para futuros estudos: recomenda-se a exploração de diversas arquiteturas para previsão espacial de séries temporais, como as redes ConvLSTM2D. Além disso, o fato de o modelo de previsão ter tido sucesso em acertar aproximadamente o local e o momento dos raios, mesmo que não acerte com exatidão o instante e nem o elemento do grid, indica que o treinamento poderia ser beneficiado caso a função de perda fosse ajustada para levar em conta não apenas acetos do ponto do grid, mas também acertos em sua vizinhança.

Referências

- Albrecht, Rachel I et al. (2016). “Where are the lightning hotspots on Earth?” Em: *Bulletin of the American Meteorological Society* 97.11, pp. 2051–2068.
- Alves, Elton Rafael et al. (2017). “Lightning prediction using satellite atmospheric sounding data and feed-forward artificial neural network”. Em: *Journal of Intelligent & Fuzzy Systems* 33.1, pp. 79–92.
- Anselmo, Evandro Moimaz (2015). “Morfologia das tempestades elétricas na América do Sul”. Tese de dout. Universidade de São Paulo.
- Ball, Geoffrey H e David J Hall (1967). “A clustering technique for summarizing multivariate data”. Em: *Behavioral science* 12.2, pp. 153–155.
- Banacos, Peter C e David M Schultz (2005). “The use of moisture flux convergence in forecasting convective initiation: Historical and operational perspectives”. Em: *Weather and Forecasting* 20.3, pp. 351–366.
- Bao, Riyang et al. (2022). “An Artificial Neural Network for Lightning Prediction Based on Atmospheric Electric Field Observations”. Em: *Remote Sensing* 14.17, p. 4131.
- Ben-Dor, Amir e Zohar Yakhini (1999). “Clustering gene expression patterns”. Em: *Proceedings of the third annual international conference on Computational molecular biology*, pp. 33–42.
- Bishop, Christopher M e Nasser M Nasrabadi (2006). *Pattern recognition and machine learning*. Vol. 4. 4. Springer.
- Bürgesser, Rodrigo E, Maria G Nicora e Eldo E Avila (2012). “Characterization of the lightning activity of “Relámpago del Catatumbo””. Em: *Journal of Atmospheric and Solar-Terrestrial Physics* 77, pp. 241–247.

- Cardoso, I et al. (2014). "Lightning casualty demographics in Brazil and their implications for safety rules". Em: *Atmospheric Research* 135, pp. 374–379.
- Cecil, Daniel J, Dennis E Buechler e Richard J Blakeslee (2014). "Gridded lightning climatology from TRMM-LIS and OTD: Dataset description". Em: *Atmospheric Research* 135, pp. 404–414.
- Cecil, Daniel J, Steven J Goodman et al. (2005). "Three years of TRMM precipitation features. Part I: Radar, radiometric, and lightning characteristics". Em: *Monthly Weather Review* 133.3, pp. 543–566.
- Chen, Suting et al. (2022). "Two-stream convolutional LSTM for precipitation nowcasting". Em: *Neural Computing and Applications* 34.16, pp. 13281–13290.
- Cheung, Mark et al. (2020). "Graph signal processing and deep learning: Convolution, pooling, and topology". Em: *IEEE Signal Processing Magazine* 37.6, pp. 139–149.
- Collins, Waylon e Philippe Tissot (2007). "2.2 USE OF AN ARTIFICIAL NEURAL NETWORK TO FORECAST THUNDERSTORM LOCATION". Em.
- Craven, Jeffrey P, Harold E Brooks, John A Hart et al. (2004). "Baseline climatology of sounding derived parameters associated with deep, moist convection". Em: *Natl. Wea. Dig* 28.1, pp. 13–24.
- Da Rocha, Rosmeri P et al. (2009). "Precipitation diurnal cycle and summer climatology assessment over South America: An evaluation of Regional Climate Model version 3 simulations". Em: *Journal of Geophysical Research: Atmospheres* 114.D10.
- Diaz, Fernando, Diego Ortiz e Francisco Roman (2022). "Lightning climatology in Colombia". Em: *Theoretical and Applied Climatology* 149.1-2, pp. 69–82.
- DiGangi, Elizabeth et al. (2022). "Analyzing lightning characteristics in central and southern South America". Em: *Electric Power Systems Research* 213, p. 108704.
- Durkee, Joshua D, Thomas L Mote e J Marshall Shepherd (2009). "The contribution of mesoscale convective complexes to rainfall across subtropical South America". Em: *Journal of Climate* 22.17, pp. 4590–4605.
- ECMWF (2017). "Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), date of access". Em.
- Fotouhi, Abbas e MJSI Montazeri-Gh (2013). "Tehran driving cycle development using the k-means clustering method". Em: *Scientia Iranica* 20.2, pp. 286–293.

- Fukushima, Kunihiko (1980). "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". Em: *Biological cybernetics* 36.4, pp. 193–202.
- Gagne, David John et al. (2017). "Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles". Em: *Weather and forecasting* 32.5, pp. 1819–1840.
- Gan, Guojun, Chaoqun Ma e Jianhong Wu (2020). *Data clustering: theory, algorithms, and applications*. SIAM.
- Garreaud, René D e John M Wallace (1997). "The diurnal march of convective cloudiness over the Americas". Em: *Monthly Weather Review* 125.12, pp. 3157–3171.
- Geerts, Bart et al. (2017). "The 2015 plains elevated convection at night field project". Em: *Bulletin of the American Meteorological Society* 98.4, pp. 767–786.
- Geng, Yangli-ao et al. (2019). "Lightnet: A dual spatiotemporal encoder network model for lightning prediction". Em: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2439–2447.
- George, J (1960). "Weather Forecasting for Aeronautics–Academic Press". Em: *New York* 41.
- Goodfellow, Ian, Yoshua Bengio e Aaron Courville (2016). *Deep learning*. MIT press.
- Goodman, Steven J et al. (2013). "The GOES-R geostationary lightning mapper (GLM)". Em: *Atmospheric research* 125, pp. 34–49.
- Groenemeijer, PH e A Van Delden (2007). "Sounding-derived parameters associated with large hail and tornadoes in the Netherlands". Em: *Atmospheric research* 83.2-4, pp. 473–487.
- Henry, W (1987). "The skew-T, Log P diagram". Em: *National Weather Service Training Center, EUA*, p. 68.
- Holle, Ronald L (2016). "A summary of recent national-scale lightning fatality studies". Em: *Weather, Climate, and Society* 8.1, pp. 35–42.
- Igel, Matthew R e Susan C van den Heever (2015). "The relative influence of environmental characteristics on tropical deep convective morphology as observed by CloudSat". Em: *Journal of Geophysical Research: Atmospheres* 120.9, pp. 4304–4322.

- Jahangir, Mohammad Hossein, Seyedeh Mahsa Mousavi Reineh e Mahnaz Abolghasemi (2019). "Spatial predication of flood zonation mapping in Kan River Basin, Iran, using artificial neural network algorithm". Em: *Weather and Climate Extremes* 25, p. 100215.
- LeCun, Yann, Yoshua Bengio et al. (1995). "Convolutional networks for images, speech, and time series". Em: *The handbook of brain theory and neural networks* 3361.10, p. 1995.
- Leinonen, Jussi, Ulrich Hamann e Urs Germann (2022). "Seamless lightning nowcasting with recurrent-convolutional deep learning". Em: *Artificial Intelligence for the Earth Systems* 1.4, e220043.
- Ma, Wenping et al. (2019). "Double-branch multi-attention mechanism network for hyperspectral image classification". Em: *Remote Sensing* 11.11, p. 1307.
- Makgatho, Ndaedzo, Ritesh Ajoodha e Hugh GP Hunt (2020). "Using Traditional Machine Learning Algorithms to Classify Lightning Occurrence". Em.
- Mandanda, Gabriel Elias (2021). *Análise dos parâmetros atmosféricos indicativos às tempestades elétricas no Sul e Sudeste do Brasil*.
- Manzato, Agostino (2013). "Hail in northeast Italy: A neural network ensemble forecast using sounding-derived indices". Em: *Weather and Forecasting* 28.1, pp. 3–28.
- Marengo, Jose A et al. (2004). "Climatology of the low-level jet east of the Andes as derived from the NCEP–NCAR reanalyses: Characteristics and temporal variability". Em: *Journal of climate* 17.12, pp. 2261–2280.
- Miller, Robert C (1975). *Notes on analysis and severe-storm forecasting procedures of the Air Force Global Weather Central*. Vol. 200. AWS.
- Moller, Alan R (2001). "Severe local storms forecasting". Em: *Severe convective storms*. Springer, pp. 433–480.
- Morales Rodriguez, CA (2019). "Thunderstorm efficiency regimes in South America as observed by STARNET and TRMM". Em: *Journal of Geophysical Research: Atmospheres* 124.21, pp. 11428–11451.
- Nobre, Carlos A, Piers J Sellers e Jagadish Shukla (1991). "Amazonian deforestation and regional climate change". Em: *Journal of climate* 4.10, pp. 957–988.
- Nyap, Lemuel Chung Chen, Waddah Waheeb e Jacqueline Luokse (2020). "Lightning prediction using fuzzy logic technique". Em: *Journal of Applied Technology and Innovation* (e-ISSN: 2600-7304) 4.3, p. 1.

- Oda, Paula SS et al. (2022). “An initial assessment of the distribution of total Flash Rate Density (FRD) in Brazil from GOES-16 Geostationary Lightning Mapper (GLM) observations”. Em: *Atmospheric Research* 270, p. 106081.
- Peixoto, José Pinto e Abraham H Oort (1992). “Physics of climate”. Em.
- Perez, Gabriel Martins Palma e Maria Assunção Faus da Silva Dias (2018). “Improving the quantitative precipitation forecast: a deep learning approach”. Em.
- Peterson, Michael (2021). “Where are the most extraordinary lightning megaflashes in the Americas?” Em: *Bulletin of the American Meteorological Society* 102.3, E660–E671.
- Rasmussen, Erik N e David O Blanchard (1998). “A baseline climatology of sounding-derived supercell and tornado forecast parameters”. Em: *Weather and forecasting* 13.4, pp. 1148–1164.
- Rasmussen, Kristen L e Robert A Houze (2011). “Orogenic convection in subtropical South America as seen by the TRMM satellite”. Em: *Monthly Weather Review* 139.8, pp. 2399–2420.
- Rasmussen, Kristen L, Manuel D Zuluaga e Robert A Houze Jr (2014). “Severe convection and lightning in subtropical South America”. Em: *Geophysical Research Letters* 41.20, pp. 7359–7366.
- Reboita, Michelle Simões et al. (2010). “Regimes de precipitação na América do Sul: uma revisão bibliográfica”. Em: *Revista brasileira de meteorologia* 25, pp. 185–204.
- “Receptive fields and functional architecture of monkey striate cortex” (1968). Em: *The Journal of physiology* 195.1, pp. 215–243.
- Reynolds, SE, M Brook e Mary Foulks Gourley (1957). “Thunderstorm charge separation”. Em: *Journal of Atmospheric Sciences* 14.5, pp. 426–436.
- Riemann-Campe, Kathrin, Klaus Fraedrich e Frank Lunkeit (2009). “Global climatology of convective available potential energy (CAPE) and convective inhibition (CIN) in ERA-40 reanalysis”. Em: *Atmospheric Research* 93.1-3, pp. 534–545.
- Romatschke, Ulrike e Robert A Houze (2010). “Extreme summer convection in South America”. Em: *Journal of Climate* 23.14, pp. 3761–3791.
- Salio, Paola, Matilde Nicolini e Edward J Zipser (2007). “Mesoscale convective systems over southeastern South America and their relationship with the South American low-level jet”. Em: *Monthly Weather Review* 135.4, pp. 1290–1309.

- Saraiva, Ivan et al. (2016). "Regional variability of rain clouds in the Amazon basin as seen by a network of weather radars". Em: *Journal of Applied Meteorology and Climatology* 55.12, pp. 2657–2675.
- Satyamurty, Prakki e Luiz Fernando De Mattos (1989). "Climatological lower tropospheric frontogenesis in the midlatitudes due to horizontal deformation and divergence". Em: *Monthly Weather Review* 117.6, pp. 1355–1364.
- Schmidt, Tobias (2023). "Gridded Hail Nowcasting using UNets, Lightning Observations, and the Warn-on-Forecast System". Em.
- Steinhaus, Hugo et al. (1956). "Sur la division des corps matériels en parties". Em: *Bull. Acad. Polon. Sci* 1.804, p. 801.
- Stolz, Douglas C, Steven A Rutledge e Jeffrey R Pierce (2015). "Simultaneous influences of thermodynamics and aerosols on deep convection and lightning in the tropics". Em: *Journal of Geophysical Research: Atmospheres* 120.12, pp. 6207–6231.
- Stolzenburg, Maribeth, W David Rust e Thomas C Marshall (1998). "Electrical structure in thunderstorm convective regions, 3, Synthesis (Paper 97JD03545)". Em: *Journal of Geophysical Research-Part D-Atmospheres-Printed Edition* 103.12, pp. 14097–14108.
- Taszarek, Mateusz et al. (2019). "A climatology of thunderstorms across Europe from a synthesis of multiple data sources". Em: *Journal of Climate* 32.6, pp. 1813–1837.
- Uccellini, Louis W e Donald R Johnson (1979). "The coupling of upper and lower tropospheric jet streaks and implications for the development of severe convective storms". Em: *Monthly Weather Review* 107.6, pp. 682–703.
- Uvo, CRB e CA Nobre (1989). "A Zona de Convergência Intertropical (ZCIT) e a precipitação no norte do Nordeste do Brasil". Em: *Boletim de Monitoramento e Análise Climática* 4, pp. 34–40.
- Valli, Márcio (2002). "Análise de cluster". Em: *Augusto Guzzo Revista Acadêmica* 4, pp. 77–87.
- Velasco, Ines e J Michael Fritsch (1987). "Mesoscale convective complexes in the Americas". Em: *Journal of Geophysical Research: Atmospheres* 92.D8, pp. 9591–9613.
- Virts, Katrina S et al. (2013). "Highlights of a new ground-based, hourly global lightning climatology". Em: *Bulletin of the American Meteorological Society* 94.9, pp. 1381–1391.

Yeditha, Pavan Kumar et al. (2020). “Forecasting of extreme flood events using different satellite precipitation products and wavelet-based machine learning methods”. Em: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30.6.

Zipser, Edward J et al. (2006). “Where are the most intense thunderstorms on Earth?” Em: *Bulletin of the American Meteorological Society* 87.8, pp. 1057–1072.